Circumventing structural uncertainty: a Bayesian perspective on nonlinear forecasting for ecology

Introduction

Ecosystems are complex, involving multiple species, age groups, and genotypes, whose interactions through time are mediated by a variety of environmental drivers. While it is possible to disentangle these sources of complexity for a handful of experimentally tractable, well-studied systems, identifying models for less studied or intractable systems is a daunting task. This is particularly relevant when models are needed to inform conservation and management decisions, where seemingly slight changes in model structure can lead to qualitatively different predictions (Lee et al., 1999; Wood and Thomas,1999).

Alternatively, nonparametric time series methods allow us to study the dynamics of a system without having to specify a model. Although this limits some of the information that can be extracted from a time series with a correctly specified model structure (e.g. estimates of relevant parameters), the insights that we can gain are robust to model misspecification.

Nonlinear forecasting developed in the 1980's and 90's based on Takens (1981) theorem of time-delay embedding. Although initially restricted to single time series from an autonomous, deterministic system, these methods have since been generalized to multiple time series (Deyle and Sugihara, 2011) from non-autonomous systems with deterministic (Starke, 1999) and stochastic forcing (Starke et al., 2003). These methods have been of great use in physics (Buzug and Pfister, 1992), neurobiology (Kannathal et al., 2005), and econometrics (Mayfield and Mizrach, 1992) where long time series that are relatively free of observation noise are fairly common.

Ecological applications of nonlinear forecasting were popular in the 1980's and 90's, including outstanding work by Sugihara (1994), Schaffer (1985), and Ellner and Turchin (1993); see Hastings et al. (1993) for a review. In the current literature, these methods seem to have been supplanted by more 'mechanistic' state-space models (see Patterson et al., 2008, and references therein) or linear models with time varying coefficients (e.g. Ives and Dakos, 2012). The primary objections to using time-delay embedding in ecology seem to be that ecological time series are noisy, not long enough to define an attractor and not stationary (Sugihara et al., 1990; Grenfell et al., 2001). Here, we address these concerns in a Bayesian nonparametric framework, using Gaussian process models to infer dynamics in delay coordinates. The advantages of the Gaussian process are its simplistic parameterization and ability to estimate with precision complicated nonlinear functions (O'Hagan,1978).

Modern statistical methods offer several advantages that are particularly relevant to ecology. First, hierarchical approaches allow us to share information across data sets without assuming that they are identical (e.g., Shi et al., 2005; Bjornstad and Grenfell, 2001; Royle and Dorazio, 2008; Halstead et al., 2012). Second, covariates can be incorporated into the dynamics in a natural and obvious way (Patterson et al., 2008). Third, nonstationary dynamics can be accommodated by allowing parameters to change with time (West and Harrison, 1997; Wikle, 2003; Ives and Dakos, 2012). Here, we present a Bayesian perspective on time-delay embedding that allows us to combine the advantages of nonparametric time series approaches with current statistical tools.

Methods and Results

To begin, we briefly describe time delay embedding and outline the basic approach of nonparametric Bayesian regression via Gaussian processes (GP). We then derive a hierarchical GP model that allows for dynamical variation among populations and extend this to the case where the dynamics are allowed to drift slowly through time. These methods are then applied to a sequence of simulated data sets.

Time-delay embedding

There is now a long history of applying Takens theorem and time delay embedding in the ecological literature, see, e.g., Schaffer (1985), Ellner and Turchin (1993), and Sugihara (1994). However, most of the descriptions of the idea are steeped in the alien vernacular of topology. From a practical point of view, the upshot of Takens theorem is that we are justified in modeling the dynamics of a single time series as $y_t = f(y_{t-1}, \ldots y_{t-L})$ for some unknown function f and 'embedding dimension' L which is at least twice the dimension of the attractor (Takens 1981). Here, a fixed time step of 1 is assumed in keeping with the majority of ecological time series applications. In settings where the data are continuously sampled in time, an appropriate time lag, Δ, must also be determined and the model is $y_t = f(y_{t-\Delta}, \ldots y_{t-L\Delta})$. Most approaches to nonlinear forecasting can be thought of as ways of approximating the unknown function f.

One particularly simple and flexible way to approximate f is using locally-weighted multiple linear regression, as in Sugihara's S-Map (Sugihara 1994). Specifically, a locally linear model of the form $y_t = \sum_{i=1}^{L} \beta_{t,i} y_{t-i} + \varepsilon_t$ is fit to the time series by least squares. We have highlighted this method in particular because it was precisely locally linear models that motivated O'Hagan (1978) to introduce Gaussian processes (GP) as priors for flexible regression modeling from a Bayesian point of view. Here, we use the tools of Bayesian GP regression to construct a hierarchical approach to nonlinear forecasting that allows integration of information from multiple time series and explicitly

deals with nonstationarity. The main text lays out the model specification and the simulations used to test each model. Further details of prior specification and posterior inference are provided in the Appendices.

1. Gaussian process time-delay embedding

Assume we have a scalar time series $y_1, \ldots y_T$, and the goal is to estimate the unknown function f that maps the history of y into the future. To simplify notation, we'll use $\mathbf{x}_\mathrm{t} = \{y_{t-1}, \ldots y_{t-L}\}$ to represent the delay-coordinate vector so that we are attempting to fit a model of the form $y_t = f(\boldsymbol{x}_t) + \epsilon_t$. The errors ϵ_t are explicitly included here to account for approximation errors as well as process noise. For convenience, we assume that ϵ_t is at least approximately normally distributed with mean 0 and variance V_ε.

The shape of f is unknown and we would like to estimate it from the available data. In a Bayesian context, we do so by assigning a prior to f and updating the distribution for f given the observed data. We are inferring a function, thus we need a prior on a space of functions and the natural place to look for these is the theory of stochastic process.

The Gaussian process is particularly convenient to work with as a prior for uncertain regression functions (O'Hagan, 1978). GP models have been used widely in spatial statistics under the guise of Kriging (Cressie, 1993). In addition they have been used in population modeling to estimate the form of density dependence (Munch et al., 2005), test for the presence of Allee effects (Sugeno and Munch, 2013), and as a tool to assess model misspecification (Thorson et al., 2014). Rasmussen and Williams (2006) is an excellent source for additional background on modeling with Gaussian processes.

The GP is a generalization of the multivariate normal distribution and is defined in terms of a mean function μ and a covariance function Σ. In the present application, $\mu = 0$. The covariance function then controls the shape of f by specifying how strongly correlated realizations of f are at different locations. In general, the stronger the correlation and the slower it decays with distance, the smoother realizations of f will be.

There are several reasonable approaches to obtaining a covariance function in the setting of time-delay embedding. In the present application we construct the covariance function as a product of correlation functions for each delay coordinate to facilitate inference of the embedding dimension and relevant lags. Specifically, the covariance between f evaluated at points $\mathbf{x}_\mathrm{t}$ and $\mathbf{x}_\mathrm{s}$ is given by $\Sigma(\mathbf{x_t}, \mathbf{x_s}) = \tau^2 \prod_{\mathrm{i}=1}^{\mathrm{L_{max}}} \mathrm{R}(\phi_i | y_{t-i} - y_{s-i}|/r)$ where τ^2controls the variance in f at a given point and R is the squared-exponential correlation function (see Appendix 1). The ϕ_i's control the 'wiggliness' of f in the i^{th} direction (time-lag) and the factor $r = \max(y) - \min(y)$ scales the delay coordinates to the unit interval. The product is taken over all lags going from 1 to $\mathrm{L_{max}}$,

which is the maximum feasible embedding dimension. In practice $\mathrm{L}_{\max}$ scales roughly as $\sqrt{T}$ (Chen and Tong, 1992) and will be less than 10 for all but the longest ecological time series.

Note that when $\phi_i = 0$, f is constant in the i^{th} direction. Thus, to facilitate identification of a parsimonious model, we used a prior on ϕ that places most weight on $\phi_i = 0$. This approach has been taken in the machine learning literature where it is referred to as ARD, automatic relevance determination, (Neal, 1996) and in the multivariate regression literature where it is referred to as the LASSO penalty (Hans, 2009). Here we use $p(\phi_i) = 2\exp[-{\phi_i}^2/\pi]/\pi$ which sets the expected number of local extrema to 1 on the unit interval (see Appendix 1). Automatic relevance determination typically uses a threshold value for ϕ_i to drop the i^{th} variable from the model, but using model selection criteria to choose among models with a range of embedding dimensions gives comparable results.

The fully specified GP model is then given by

$$p[y_t|f, \boldsymbol{x}_t, V_\epsilon] \sim N(f(\boldsymbol{x}_t), V_\epsilon)$$

$$p[f|\tau^2, \boldsymbol{\phi}] \sim GP(0, \Sigma)$$

$$p[V_\epsilon, \tau^2, \boldsymbol{\phi}] \qquad (1)$$

The final line represents prior specification for the variance and length scale parameters. These are detailed in Appendix 1.

Full Bayesian inference for f and the hyperparameters can be obtained via MCMC or other methods (Rasmussen and Williams, 2006). However such a computationally intensive approach is impractical for simulation studies. Instead, we fix the hyperparameters at the MAP, maximum *a posteriori*, estimates (Rasmussen and Williams, 2006). To find the posterior mode, we use the R-prop algorithm developed for neural networks (Riedmiller and Braun, 1993) because it is more stable for training GP models than other numerical optimizers (Blum and Reidmiller, 2013; Nocedal and Wright, 1999). Given the MAP estimates of the hyperparameters, the posterior for f is also a GP, $f|\boldsymbol{x}, \mathbf{y}, \{\tau^2, \boldsymbol{\phi}, V_\epsilon\}_{\mathrm{MAP}} \sim GP(m_c, \Sigma_c)$ where m_c and Σ_c are the posterior mean and covariance functions obtained using standard formulae for conditioning in multivariate normals (See Appendix 2).

In the context of dynamical modeling, it is often the case that we want to know what the equilibria are and to characterize their stability. Since f is uncertain we need to think instead about the distribution of plausible equilibria. Specifically, we seek the probability, $p(x^*)$, that the state x^* is a fixed point (i.e., $x^* = f(x^*, x^*, \ldots)$) and obtain this by evaluating the posterior density along the line where $x = f$. Given the

hyperparameters, the distribution for f at the point x is just a normal density, hence $p(x^*)$ is simply

$$p(x^*) = \frac{q}{\sqrt{2\pi\Sigma_c(x^*,x^*)}} \exp\{-\frac{1}{\Sigma_c(x^*,x^*)}[x^* - m_c(x^*)]^2\} \quad (2)$$

where q is the normalization constant. Note that this is not a normal distribution in x^* because both m_c and Σ_c depend on x^*. This generalizes immediately to higher dimensions by evaluating $m_c(x^*, x^*, \ldots)$ and $\Sigma_c(x^*, x^*, \ldots)$.

To test the utility of the GP framework for identifying the embedding dimension, relevant lags, and plausible equilibria, we simulated data from models with $n_t = rn_{t-1}\exp[-n_{t-d} + \epsilon_t]$ where d ranged from 1 to 4 with corresponding values of $r = [8, 3.5, 2.5, 1.75]$ and $\epsilon_t \sim N(0, 0.1)$. Importantly, the relevant lags are spelled out in the model formulation, making the evaluation of ARD lag selection totally unambiguous. This is in contrast to simulating a sequence of increasingly complex multi-species models, which provide no *a priori* way to determine which lags are most relevant except in the simplest cases. For each value of s, we simulated 1000 time series of length 100. For each simulated data set, we fit the GP model with $L_{max} = 6$ and used $\phi_i > 0.1$ as the threshold to determine if the i^{th} lag is relevant in the model. We then computed the distribution of plausible equilibria and selected the posterior mode as the best estimate for that data set.

Overall, the ARD approach seems to identify the relevant lags correctly a majority of the time (Figure 1). The chief exception to this is when $d = 3$, where lags 2 and 3 appear relevant with roughly equal frequency. The MAP estimates of the plausible equilibria are also generally quite close to the true value obtained from the deterministic skeleton (Figure 1).

2. Hierarchical modeling

The most compelling reasons to formulate a Bayesian approach to time-delay embedding are the extensions that it invites. Here we extend the GP time-delay embedding model to accommodate multiple short time series. Previously, Hsieh et al. (2008) demonstrated that multiple short series can be concatenated to improve forecasts. However, this requires that the attractor is the same for each series. Although we expect underlying similarities across series to be present (e.g. for the same species in several locations), we also expect population-specific features to exist. In

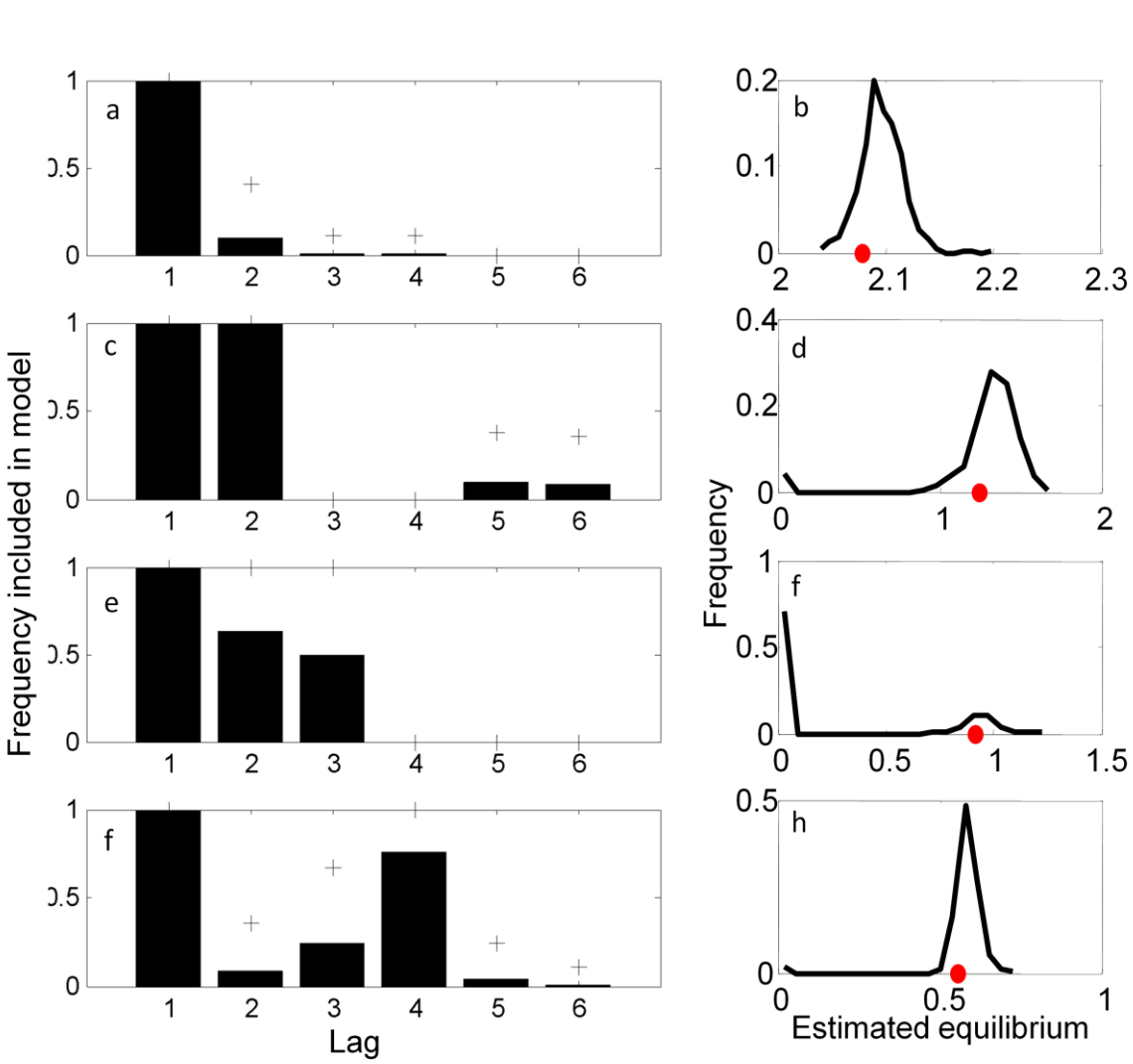


Figure 1. Automatic relevance determination and putative equilibria. From top to bottom, rows present results for $d = 1$ to $d = 4$. For each simulation, lag 1 is always relevant. Left: Bars indicate the frequency with which each lag coordinate was relevant in the fitted model (i.e. the number of simulations in which $\phi_i < 0.1$). Right: For each simulation, the posterior for plausible equilibria was calculated and the value with the maximum probability was recorded. Black lines show the distribution of these estimates across 1000 simulated data sets while the red dot is the true value for a fixed point for the deterministic skeleton.

order to account for this, we propose a hierarchical model structure that allows the degree of similarity between series to be determined from the data.

In the standard approach to hierarchical Bayesian models, information is shared across data sets by asserting that their parameters come from a common distribution (e.g., Shi et al., 2005; Bjornstad and Grenfell, 2001; Royle and Dorazio, 2008; Halstead et al.,

2012). For example, we might improve estimation of the mean for under sampled populations by treating each population-specific mean as a random deviation from the cross-population mean. In the present case we do something analogous, by proposing that $f_i = \mu + z_i$ where $z_i \sim GP(0, \Sigma)$ and assigning μ a $GP(0, C)$ prior. The covariance function C is defined analogously to Σ in (1) but with potentially different point-wise variance and length scale parameters, i.e., σ^2 and $\boldsymbol{\gamma}$ respectively. Marginalizing over μ we obtain the total point-wise variance in f given by $\varpi^2 = \tau^2 + \sigma^2$. The correlation between maps for two different populations at the same state is $Corr[f_i(\boldsymbol{x}), f_j(\boldsymbol{x})] = C(\boldsymbol{x}, \boldsymbol{x})/[C(\boldsymbol{x}, \boldsymbol{x}) + \Sigma(\boldsymbol{x}, \boldsymbol{x})]$. If we write $\sigma^2 = \rho\varpi^2$ and $\tau^2 = (1-\rho)\varpi^2$ with ρ between 0 and 1, the correlation between maps reduces to ρ. This partitions the variance into within- and across- series components, such that all f's are identical when $\rho = 1$ and independent when $\rho = 0$. In this way, information is shared across data sets without assuming that the dynamics are identical. The GP model for this case is given by

$$p[y_{it}|f_i, \boldsymbol{x}_{it}, V_\epsilon] \sim N(f_i(\boldsymbol{x}_t), V_{\epsilon i})$$

$$p[f_i|\boldsymbol{\theta}] \sim GP(\mu, \Sigma)$$

$$p[\mu|\boldsymbol{\theta}] \sim GP(0, C)$$

$$p[\boldsymbol{V_\epsilon}, \boldsymbol{\theta}] \qquad (3)$$

where $V_{\epsilon i}$ is the population specific process variance and $\boldsymbol{\theta}$ collects all of the other parameters. In this model we set a uniform prior, $U(0,1)$, on ρ.

To demonstrate the utility of this framework we simulated data from 12 models with $1-5$ states (see Table 1) and repeated these over 3 independent locations. These models represent a wide range of ecological phenomena including delayed regulation (e.g. Turchin 1990), seasonal variation in productivity (e.g. Summers et al 2000), density-dependent maturation (Neubert and Caswell 2000), maternal effects (Ginzburg and Taneyhill, 1994), host-parasitoid interactions (Beddington et al. 1975), competition (e.g. Schoombie and Getz, 1998), contemporary evolution (Doebeli & deJong, 1999), and migration (e.g. Gerber et al., 2002). When the models include multiple classes of individuals, we use time series from the same focal class for each location. Each model was simulated under 3 baseline parameter sets that generate fixed point, limit cycle, and chaotic dynamics respectively. Among populations, model parameters were allowed to vary from this baseline between 0 and 25%. In these simulations, we focused on short time series and evaluated the improvement in the out of sample forecast precision for series of 10, 15, and 20 points. To quantify forecast precision, we produced 1-step ahead forecasts for the final 5 years of data for each location and computed the total forecast error as $SS = \sum_{i=1}^{3}\sum_{t=T-5}^{T}[y_{it} - f_i(\boldsymbol{x}_{it})]^2$. To do so, we computed the MAP estimates of $\boldsymbol{\theta}$ based on the first $T-6$ years of data, and

conditioned on these to produce estimates for $T-5$ to T. For comparison, we repeated this calculation with each series modeled independently.

Figure 2 illustrates the improvement over independent modeling for the density dependent maturation model. Overall, the hierarchical approach reduced forecast error in series of length 10 by $\sim 60\%$ on average, but this drops off quickly as T increases. This is consistent with the observation that low dimensional dynamics (such as we simulated) can typically be reconstructed with $30-40$ points. Increasing process variance increased the fraction of simulations in which the hierarchical model outperformed the independent model, but diminished the average magnitude of error reduction. Interestingly, neither the dynamical regime nor the variance among populations had as strong an effect on the relative performance of the hierarchical model.

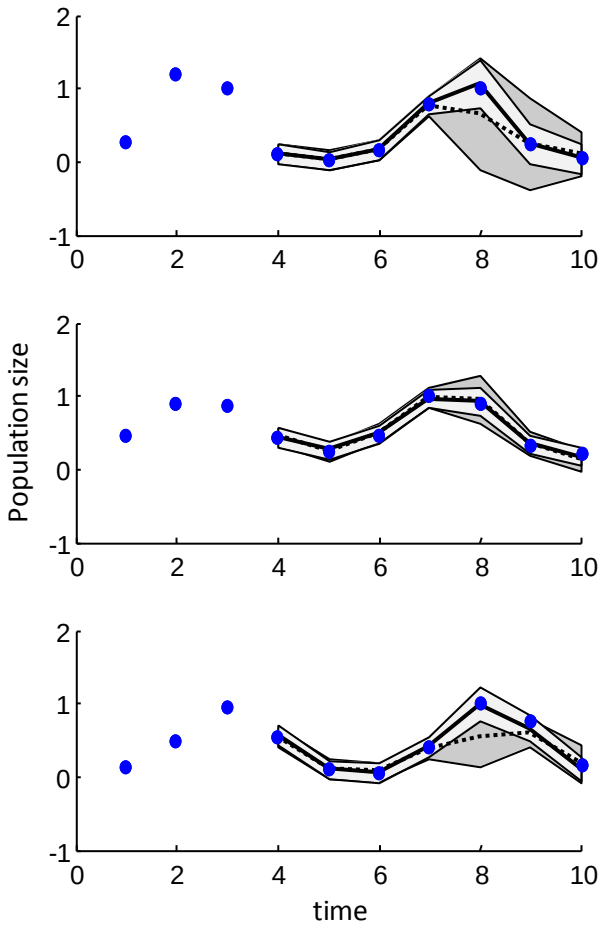


Figure 2. An illustration of hierarchical GP forecasting. The time series are generated by the density-dependent maturation model (Table 1). Blue points in each panel represent time series for 3 independent realizations of the model with 10% variation among parameters. The solid black line (mean) and light grey areas (+/- 2 sd) indicate 1-step ahead forecast from the hierarchical GP model. The dashed line and dark grey interval indicate results for each population modeled independently.

Table 1: Simulation models used to evaluate the GP embedding approach. For each model one parameter is varied to produce chaotic, limit cycle, and stable dynamics. For all models, except the delayed logistic, the noise term $\boldsymbol{z_t}$ is drawn independently from $N(-s^2/2, s^2)$. Noise in the delayed logistic model must respect the requirement that x remain within [0,1] at all time, thus, we set $w_t \sim U[s(1-N_{t-\tau}), s(N_{t-\tau}-1)]$. In several of the models the noise term follows a '$\circ$', which indicates elementwise multiplication of the noise vector with the deterministic next-state vector.

Scenario	Model	Parameter Values
Delayed regulation (Logistic)	$N_{t+1} = rN_t(1+N_{t-\tau})(1+w_t)$	$\tau = 1$ $r = \{2.1, 2.19, 1.54\}$
Delayed regulation (Hassel)	$N_{t+1} = rN_t(1+N_{t-\tau})^{-b}e^{z_t}$	$r = 4, \tau = 1$ $b = \{3, 2, 1.75\}$
Seasonal productivity	$N_{t+1} = N_t e^{r[1+\alpha \sin(2\pi t/\theta)]-N_t+z_t}$	$r = 1.95, a = 3$ $\theta = \{4, 3, 8\}$
Density-dependent maturation	$\begin{bmatrix}A\\J\end{bmatrix}_{t+1} = \begin{bmatrix}s_A & s_J g(A_t+J_t)\\ b\mathrm{e}^{z_t} & s_J[1-g(A_t+J_t)]\end{bmatrix}\begin{bmatrix}A\\J\end{bmatrix}_t$ $g(x) = G_{max}\exp(-\gamma x)$	$s_a = 0.1, s_j = 0.5$ $G_{max} = 0.9,$ $\gamma = 0.01$ $b = \{35, 34, 24\}$
Host-parasitoid	$N_{t+1} = rN_t e^{-N_t-\gamma P_t+z_t}$ $P_{t+1} = \alpha N_t e^{-N_t+z_t}(1-e^{-\gamma P_t})$	$\gamma = 0.5, \alpha = 2$ $r = \{7, 8, 5\}$
Competition (Hassell-Comins)	$N_{1,t+1} = r_1 N_{1,t} e^{z_{1,t}}[1+(N_{1,t}+cN_{2,t})/\gamma]^{-\gamma}$ $N_{2,t+1} = r_2 N_{2,t} e^{z_{2,t}}[1+(N_{2,t}+dN_{1,t})/\gamma]^{-\gamma}$	$c = 0.9, d = 0.8$ $\gamma = 20, r_1 = 1.5r_2$ $r_2 = \{16, 12, 5\}$
Competition (Shepherd)	$N_{1,t+1} = r_1 N_{1,t} e^{z_{1,t}}/\{1+(r_1-1)[N_{1,t}+cN_{2,t})]^{\gamma}\}$ $N_{2,t+1} = r_2 N_{2,t} e^{z_{2,t}}/\{1+(r_2-1)[N_{2,t}+dN_{1,t})]^{\gamma}$	$c = 0.2, d = 0.1$ $\gamma = 4, r_1 = 1.5r_2$ $r_2 = \{3, 2.4, 1.6\}$
Contemporary evolution	$\begin{bmatrix}N_{AA}\\N_{Aa}\\N_{aa}\end{bmatrix}_{t+1} = \begin{bmatrix}p_{AA}+p_{Aa} & p_{Aa}/4 & 0\\ p_{aa} & p_{Aa}/2+p_{aa} & p_{AA}\\ 0 & p_{Aa}/4 & p_{AA}+p_{Aa}\end{bmatrix}\begin{bmatrix}N_{AA}/(1+ax^c)\\N_{Aa}/(1+bx^c)\\N_{aa}/(1+ax^c)\end{bmatrix} \circ e^{\boldsymbol{z}_t}$	$a = 4*8^{-c},$ $b = 0.9\mathrm{a}$ $c = \{4, 3.35, 2.5\}$

$$x = \sum N\,, p_{ij} = N_{ij}/x$$

Table 1. *Continued. Simulation models with migration. In the 5-location migration models, the migration matrix,* $\boldsymbol{M}$*, represents the exchange of individuals across sites. We used 3 different migration topologies representing a linear chain with reflective boundaries, a ring, and uniform dispersal. In each case the parameter* μ' *is the fraction of the resident population that does not migrate and the* $(1-\mu')$ *migrants are distributed equally over all accessible neighbors. The same parameters are used for all 5 sites. The noise term* $\boldsymbol{z_t}$ *is drawn independently from* $N(-s^2/2, s^2)$*for each site.*

Scenario	Model	Parameter Values
Migration (2 locations, Ricker)	$\begin{bmatrix} N_1 \\ N_2 \end{bmatrix}_{t+1} = \begin{bmatrix} m_{11} & (1-m_{22}) \\ (1-m_{11}) & m_{22} \end{bmatrix} \begin{bmatrix} N_1 e^{r_1 - N_1} \\ N_2 e^{r_2 - N_2} \end{bmatrix} \circ e^{\boldsymbol{z}_t}$	$m_{11} = 0.35$, $m_{22} = 0.25$ $r_2 = 18$, $r_1 = \{15, 13, 7\}$
Migration (5 locations, Shepherd)	$\begin{bmatrix} N_1 \\ \vdots \\ N_5 \end{bmatrix}_{t+1} = \boldsymbol{M}(\mu') \begin{bmatrix} rN_{1,t}/\{1+(r-1)N_{1,t}{}^{\gamma}\} \\ \vdots \\ rN_{5,t}/\{1+(r-1)N_{5,t}{}^{\gamma}\} \end{bmatrix}_t \circ e^{\boldsymbol{z}_t}$	$\mu' = 0.9, \gamma = 4$ $r = \{2, 1.3, 1.2\}$ M – migration matrix, see caption

Table 2. Hierarchical GP performance. Error Red. is the proportional error reduction calculated as 1-mean(SS_{hier}*)/mean(*SS_{ind}*) and Freq. Imp. is the fraction of simulations in which the forecast error for the hierarchical model was less than for series modeled independently. These are reported for the main effects of dynamical regime, process variance, variation among populations, and the length of the time series. Performance metrics reported for each main effect are averaged over all models and other parameters.*

	Dynamical Regime			*Process Variance*		
	Chaos	Limit Cycle	Stable	0	0.05	0.1
Error Red.	0.57	0.56	0.41	0.68	0.56	0.44
Freq. Imp.	0.84	0.81	0.73	0.76	0.81	0.81
	Variation among populations			*Length of series*		
	0	0.1	0.25	10	15	20
Error Red.	0.56	0.54	0.47	0.61	0.33	0.22
Freq. Imp.	0.82	0.8	0.75	0.87	0.78	0.73

3. Dynamic embedding for nonstationarity.

It has long been recognized that for time-delay embedding to work well, long series are needed. Unfortunately, the longer the time period over which ecological data are collected, the more likely it is that some aspect of the system will change. Changes in climate, species introductions, and changes in management practices are common and recent studies suggest that nonstationary dynamics are present in a broad range of systems, e.g., multidecadal shifts in the dynamics of pacific ecosystems (Hare and Mantua, 2000). When using mechanistic models, a common approach to compensate for the apparent nonstationarity in a time series is to allow the parameters of the models to vary with time, either in a specified way or by letting them drift randomly (West and Harrison, 1997; Wikle, 2003; Ives and Dakos, 2012). This concept is the basis of the Dynamic Linear Model (DLM, West and Harrison, 1997) which is the framework used to identify regime shifts (Carpenter, 2003). In DLMs, the parameters of the model can change over time as new data is observed, allowing the model to adapt to changes in the underlying ecosystem dynamics. The framework of GP time-delay embedding can be readily extended in a similar fashion to accommodate nonlinear changes in the system. More specifically, rather than letting the parameters of the system change over time, we allow f to vary. We do this by setting $f_{t+1} = f_t + \zeta_t$, where $\zeta_t \sim GP(0, W)$ and the prior for f at time 0 is $f_0 \sim GP(0, \Sigma)$.

Rather than specify an independent covariance function for the drift variance, we set $W = \delta\Sigma$ where the scalar δ ranges from 0 to 1 and controls the rate at which f, and therefore the relationship between y_t and $\boldsymbol{x}_t$ changes as time progresses. We note that as $\delta \to 0$, f changes less through time and when $\delta = 0$ the model reduces to the basic GP time-delay embedding model (section 1). We can gain some intuition for how f changes through time under this prior specification by looking at the correlation between f_t and f_0: $Corr(f_t(\boldsymbol{x_t}), f_0(\boldsymbol{x_s})) = (1+\delta t)^{-1/2} \prod_{i=1}^{L_{max}} R(\gamma_i | y_{t-i} - y_{s-i}| / r)$. At a specific state $\boldsymbol{x}$, the correlation decays with t as$(1+\delta t)^{-1/2}$ becoming effectively independent for long t. For a single population, this 'dynamic embedding' model is given by

$$p[y_t | f_t, \boldsymbol{x}_t, V_\epsilon] \sim N(f_t(\boldsymbol{x}_t), V_\epsilon)$$

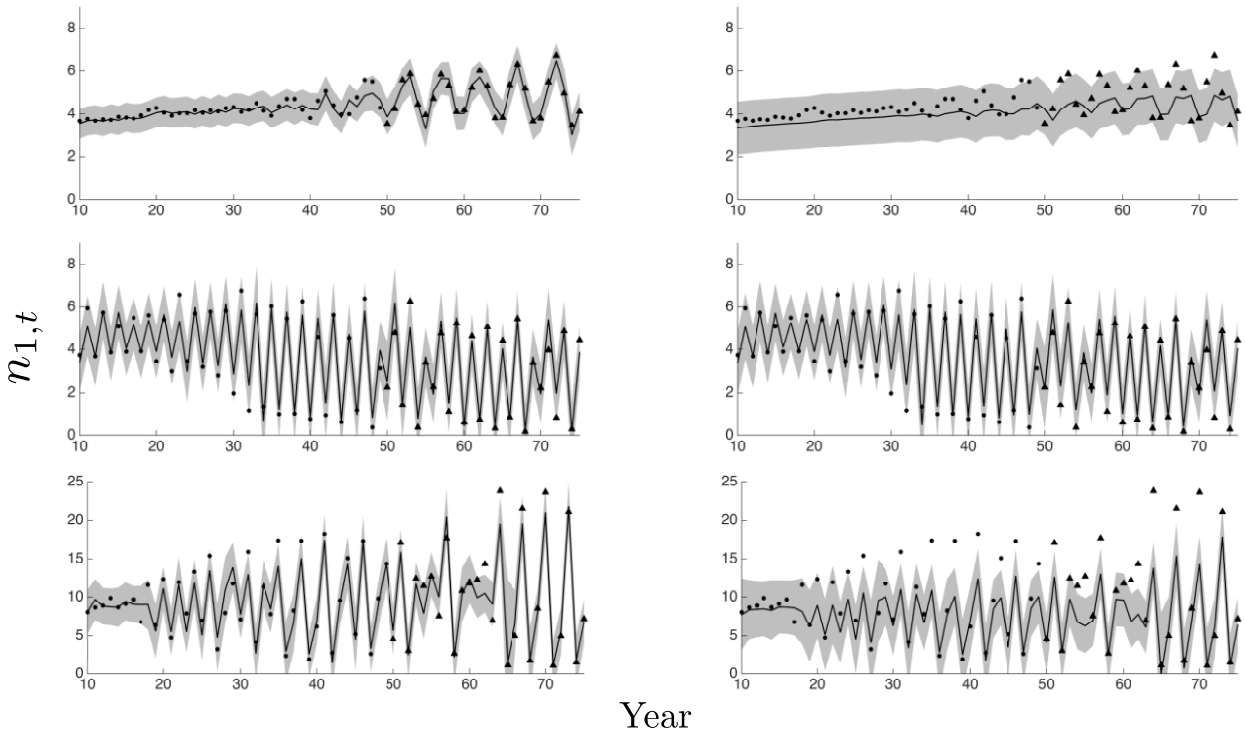
$$p[f_{t+1} | f_t, \boldsymbol{\theta}] \sim GP(f_t, \delta\Sigma)$$

$$p[f_0 | \boldsymbol{\theta}] \sim GP(0, \Sigma)$$

$$p[V_\epsilon, \boldsymbol{\theta}] \qquad (4)$$

In this model, we assign $1/\delta$ an exponential prior with mean $0.49/T$ so that the expected time for the correlation to drop to 0.1 is $\sim 2T$.

To evaluate the performance of the dynamic embedding model, we simulate data from three of the models described in section 2: density-dependent maturation, migration (2 location, Ricker), and maternal effects. In this section, the goal is to determine our ability to forecast nonstationary dynamics; therefore, we allow the growth rate, migration rate and the maximum reproductive rate to vary through time in each of the models respectively (see Table 3). We simulate 75 years of data from each of the models, use the first 50 years to determine the model parameters and compute one-step predictions for the last 25 years conditional on the MAP estimates. We compare the forecast error



of the dynamic embedding model against that of time-constant model as in section 2.

Figure 3: Performance of the dynamic embedding model (left) relative to one without drift (right). The black line (mean) and grey patch (+/- 2 sd) illustrate one-step-ahead predictions based on data (black dots) from three different ecological scenarios (Top row: Density dependent maturation, center row: Two location Ricker, bottom row: Maternal effects). In each case 1 parameter of the simulation model is changing through time. The stationary GP compensates for this by making the noise variance high and the length scale short, which tends to produce imprecise forecasts. In contrast, the dynamic embedding model is robust to slow changes in the underlying dynamics.

Although the attractor changes shape as the driving variable changes, the standard GP does a fair job of forecasting, but does so by inflating the estimated process noise (Figure 3). In contrast, the dynamic embedding model produces more accurate one-step forecasts with smaller confidence bands.

Overall, the dynamic embedding model reduces forecast error for most simulations (Table 3). The magnitude of the improvement is 20-30% for most of the small-noise scenarios ($s^2 = 0.05$ and 0.1). For $s^2 = 0.2$, the error reductions are typically 10-20%. The main exception to this is for parameter set 1 of the migration model. For this model the time-constant GP seems to do quite well despite the time-varying parameter. Thus, adding drift actually reduces model performance.

Table 3. Dynamic embedding results for three ecological scenarios. For each ecological scenario, time series were simulated with one parameter (Driver) increasing linearly through time over the range indicated. This was repeated under three different parameter sets (see table 1 for model definitions and parameter values) and three different levels of process stochasticity. Because of the driving variable, the dynamical regimes for the 3 parameter sets from Table 1 are no longer strictly chaotic, limit cycle, and stable. For each combination, we report the proportional reduction in 1-step forecast error for the dynamic embedding model compared to one without temporal drift. The numbers in parentheses represent the percentage of simulations that the dynamic embedding model outperformed the independent model.

Model Name	Driver	Set	$s^2 = 0.05$	$s^2 = 0.1$	$s^2 = 0.2$
Density dependent maturation	$G_{max} \in (0.25,0.99)$	1	0.2898 (100%)	0.3748 (100%)	0.2312 (100%)
		2	0.3365 (100%)	0.3343 (100%)	0.2337 (100%)
		3	0.2064 (100%)	0.2086 (100%)	0.1922 (100%)
Migration (2 locations, Ricker)	$m_1 \in (0.25,0.8)$	1	-0.0252 (31.4%)	0.0422 (82.9%)	0.0293 (71.4%)
		2	0.0687 (80.0%)	0.0371 (74.3%)	0.0227 (74.3%)
		3	0.1223 (91.4%)	0.1332 (97.1%)	0.0351 (80.0%)
Maternal Effects	$r \in (1.5,5.2)$	1	0.4471 (100%)	0.3256 (100%)	0.1228 (97.1%)
		2	0.4283 (100%)	0.3205 (100%)	0.1271 (100%)
		3	0.4381 (100%)	0.3204 (100%)	0.1080 (88.6%)

Discussion

Here we have developed Bayesian approaches to time delay embedding for use in ecological forecasting. The Bayesian paradigm offers a number of advantages over algorithmic approaches including automatic quantification of uncertainty, incorporation of prior information, and detection of relevant lags. However, the biggest advantage is likely to be the ease with which the models are generalized. The hierarchical and dynamic embedding models we presented were developed to cope with the short, noisy, and nonstationary time series encountered in ecology. As shown in Appendix 3,

combining these models is trivial, though we do not present simulation results along these lines.

The simulations presented here are promising and demonstrate good performance over a broad range of scenarios. But they are certainly not exhaustive. Our primary intent is to demonstrate proof of concept. Having done so, it would be worthwhile to determine the limits under which the models fail. For instance, although the hierarchical model reduces error when the simulation parameters differ by up to 25%, there must surely be a limiting difference beyond which the model is not useful. Similarly, we expect the dynamic embedding model to be useless when the driving variable changes sufficiently fast.

Combining short time series through hierarchical GP models increases forecast precision across a wide range of ecological scenarios. In cases where more than two or three series are available, it may be worthwhile to generalize the hierarchical structure to allow different degrees of dependence across pairs. Rather than have the pointwise correlation be constant for all pairs, ρ, we could instead model pairwise correlations separately, e.g., $Corr(f_i(\boldsymbol{x}_t), f_j(\boldsymbol{x}_t)) = \rho_{ij}$ for the i^{th} and j^{th} populations. This approach offers a means of identifying clusters of series with similar dynamics. If the dynamics are expected to vary along spatial gradients, a natural choice would be to use a spatial covariance function to constrain the correlations across embeddings.

We have focused primarily on the situation where multiple series are available for a single species. For any given location it is often the case that data are available for more than one species. In this case, the most direct thing to do is to fit GP models for each species using the data for all species as predictors, i.e. we could write $x_t = f(x_{t-1}, y_{t-1}, \ldots) + \varepsilon_t$. This could be done easily using the methods we have described with little modification. However, it is worth noting that the multivariate embedding theorem (Deyle and Sugihara, 2011) tells us that we are justified in using any combination of E lags from all interacting species. This suggests that we might improve forecast precision using model averaging to combine multiple embeddings.

The dynamic embedding approach we propose for handling nonstationary series substantially improves our ability to forecast when the underlying dynamics are changing. In addition to allowing us to forecast in the presence of nonstationarity, we suspect that the dynamic embedding approach could be of use in anticipating ecological regime shifts. Specifically, we might compare the fit of models with and without temporal drift to test for the presence of nonstationarity.

Our approach to nonstationary dynamics assumes that we know little beyond the fact that the system is changing. If we had some information on *how* the system was changing there are several obvious alternatives. For instance, if the driving variable is

actually known, we could simply include it as another state in the GP model. Alternatively, if the driver is unknown, we could write the dynamics in terms of a single extra variable that drifts through time, rather than allowing the shape of the inferred map to change. That is, we could write $x_t = f(x_{t-1}, \ldots, x_{t-E}, s_t) + \varepsilon_t$ with $s_t = s_{t-1} + w_t$. It may be that a model with a low-dimensional drift term is more efficient than the infinite-dimensional drift model we have proposed.

All of the models we have proposed and all of the simulations we have done have ignored observation uncertainty. This assumption makes it possible to analytically marginalize over the unknown f, simplifying the problem down to estimating a handful of hyperparameters. With observation errors, this is no longer the case and we must resort to either MCMC (Hastings, 1970; Ming-Hui et al., 2000) or Laplace approximation (Tierney and Kadane, 1986). Accounting for observation errors in GP-based dynamical models is certainly possible (Thorson et al., 2014), though at the cost of dramatically increased computation. It is also worth noting that in simulations using Sugihara's S-map for forecasting, the results are robust to modest levels of measurement error (Perretti et al., 2013; and Deyle et al., 2013). That said, extending the models shown here to state-space settings is an important area for future development.

Nonlinear dynamics and chaos have been clearly demonstrated in multiple experimental systems (Ellner and Turchin, 1993; Desharnais et al., 2001; Becks and Arndt, 2008), and are likely to be the norm in ecology. Similarly, structural uncertainty is a pervasive problem in ecological modeling and ecosystem management. The tools we propose here do not offer a solution to the problem of structural uncertainty. Rather, they give us a way to avoid the problem entirely - when the goal is to generate short term predictions and make robust management decisions. We anticipate that these methods will be of value in conservation and management whenever the 'true' dynamics of the system are unknown.

Acknowledgements

This work was supported by funding from the Lenfest Ocean Program and NOAA's IAM program. It has greatly benefitted from input by George Sugihara, Ethan Deyle, Hao Ye, Athansios Kottas, Alec MacCall, and Marc Mangel. S Munch is grateful to MBI, Alan Hastings, Jennifer Dunne, and Andrew Morozov for organizing the symposium on Uncertainty, Sensitivity, and Predictability in Ecology.

Appendix 1. *Prior specification*

We set $p(\phi)$ such that the expected number of local extrema is 1, in keeping with our intuition that the function we are attempting to estimate should not be too 'wiggly.' The scale factor $r = [max(y) - min(y)]$ has been included so that the ϕ_i's are dimensionless and the range of $|y_{t-i} - y_{s-i}|/r$ is $[0,1]$. We can then make use of the fact that the expected number of zero crossings of a stationary GP on the unit interval (Sacks and Ylvisaker, 1966) is given by $E(\#\text{ of crossings in } [0,1]) = \pi^{-1}[-\mathrm{R}''(0)]^{1/2}$. Combining this with the fact that the derivative of a GP is also a GP with covariance function given by $\partial^2\Sigma/\partial y_t \partial y_s$ (see Rasmussen and Williams, 2006), the expected number of turns (local extrema) on the unit interval is given by $\mathrm{E}(\#\text{ turns in } [0,1]) = \pi^{-1}[\mathrm{R}^{(4)}(0)]^{1/2}$.

Using correlation function $R = exp\,[-(\phi d)^2]$ with $d = |y_t - y_s|/r$, we have $R^{(4)}(0) = 12\phi^4$ so that $E(\#\text{ turns in } [0,1]|\phi) = \pi^{-1}\sqrt{12}\phi^2$. Since we want $E(\#\text{ turns in } [0,1]) = \pi^{-1}\sqrt{12}\phi^2 \approx 1$, we set our prior for ϕ so that this is true.

We used weakly informative priors for the remaining parameters, τ^2 and V_ε. Specifically, we set $P(\tau^2/\mathrm{Var(y)}) = P(\mathrm{V_\varepsilon}/\mathrm{Var(y)}) = \mathrm{Beta}(1.1, 1.1)$ such that the prior mean for each parameter is $\mathrm{Var}(y)/2$. Although this may seem restrictive, this prior only constrains the total variance in the predicted population size, e.g, y_{T+1}, to be less than twice the observed variance in $[y_1, \ldots, y_T]$.

Appendix 2. *Updating*

For all of the cases described in the main text, updating uses a two-stage approach. In the first stage, the parameters and hyperparameters $\{V_\epsilon, \tau^2, \boldsymbol{\phi}$, etc.$\}$ are estimated from the marginal posterior obtained by integrating out f. Sampling for these parameters is done using a Metropolis Hasting algorithm. In the second stage, we make use of the fact that the posterior distribution for f given the parameters is also a GP

$$p[f|\boldsymbol{y}, \tau^2, \boldsymbol{\phi}] \sim GP(m_c, \Sigma_c)$$

where the conditional mean and covariance functions, evaluated at some new states $\boldsymbol{x_{new}}$ are given by

$$m_c(\boldsymbol{x}_{new}) = \Sigma(\boldsymbol{x}_{new}, \boldsymbol{X})[\Sigma(\boldsymbol{X}, \boldsymbol{X^T}) + V_\epsilon \boldsymbol{I}]^{-1}\boldsymbol{y}$$

$$\Sigma_c(\boldsymbol{x}_{new}, \boldsymbol{x'}_{new}) = \Sigma(\boldsymbol{x}_{new}, \boldsymbol{x'}_{new}) - \Sigma(\boldsymbol{x}_{new}, \boldsymbol{X})[\Sigma(\boldsymbol{X}, \boldsymbol{X^T}) + V_\epsilon \boldsymbol{I}]^{-1}\Sigma(\boldsymbol{X^T}, \boldsymbol{x'}_{new})$$

where $\mathbf{X}$ is the collection of delay vectors for the observed series $\{\boldsymbol{x}_L, \ldots, \boldsymbol{x}_T\}$ and $\boldsymbol{y}$ is the vector of observed next states $\{y_L, \ldots, y_T\}$.

Using this updating scheme, we can produce 1-step ahead forecasts directly by setting $\boldsymbol{x}_{new} = \boldsymbol{x}_{T+1}$. If we are interested in N-step forecasts for $N > 1$, two approaches are possible. The first approach is to construct a distribution of future states by simply iterating the estimated f over several steps. A second approach, is to estimate a new function, say f_N that produces an N-step forecast directly, by replace the vector $\boldsymbol{y} = \{y_L, \ldots, y_T\}$ with $\boldsymbol{y_N} = \{y_{L+N}, \ldots, y_{T+N}\}$. Here we adopt the latter approach as it is considerably faster and less prone to spurious error inflation when N is not large and time series used to fit the GP is reasonably long.

Appendix 3. *Hierarchical models with time-varying dynamics.*

The hierarchical and nonstationary models can be readily combined, allowing the population-specific f's and the across-population mean μ to drift through time. As in the previous two cases there is an underlying additive representation for f, which is now $f_{it} = \mu_t + z_{it}$ and $\mu_{t+1} = \mu_t + \eta_t$ with the addition of a random walk component for the deviations from the mean, i.e. $z_{it+1} = z_{it} + \zeta_{it}$. The intial conditions are all represented by GP's: $\mu_0 \sim GP(0, C)$ and $z_{i0} \sim GP(0, \Sigma)$ and discount factors are used to parameterize the temporal variations: $\eta_t \sim GP(0, \delta C)$ and $\zeta_{it} \sim GP(0, \psi\Sigma)$. The discount factor δ controls the rate at which μ changes through time, while the new discount factor ψ controls the rate at which specific populations drift independently. As in the hierarchical case, we set the point-wise variance for μ_0 to $\rho\varpi^2$ and the pointwise variance in z_{i0} to $(1-\rho)\varpi^2$. The two previous models are obtained as special cases of this model by setting $\psi = \delta = 0$ for the hierarchical case or setting $\rho = 0$ to obtain independent time varying models for each population.

For building intuition, it is again useful to think about the correlation between two f's at a single state $\boldsymbol{x}$ and time t. Across populations, we have

$$Corr(f_{jt}(\boldsymbol{x}), f_{kt}(\boldsymbol{x})) = \frac{\rho(1+\delta t)}{\rho(1+\delta t) + (1-\rho)(1+\psi t)}$$

If the drift rates are equal, $\psi = \delta$, then the correlation remains ρ for all t. If the series do not drift independently ($\psi = 0$) , then the correlation goes to 1 and if $\delta = 0$ then the correlation goes to 0.

The correlation through time for a single population is given by

$$Corr(f_{j0}(\boldsymbol{x}), f_{jt}(\boldsymbol{x})) = \frac{1}{\sqrt{\rho(1+\delta t) + (1-\rho)(1+\psi t)}}$$

The introduction of z as an additional GP was just a notational convenience. Since $z_{it} = f_{it} - \mu_t$, we can eliminate it from the model. Doing so, the hierarchical model with time varying dynamics is given by

$$p[y_{it}|f_{it}, \boldsymbol{x}_{it}, V_{\epsilon i}] \sim N(f_{it}(\boldsymbol{x}_{it}), V_{\epsilon i})$$

$$p[f_{i,t}|\mu_t, f_{i,t-1}, \mu_{t-1}, \boldsymbol{\theta}] \sim GP(\mu_t + f_{i,t-1} - \mu_{t-1}, \psi\Sigma)$$

$$p[\mu_{t+1}|\mu_t, \boldsymbol{\theta}] \sim GP(\mu_t, \delta C)$$

$$p[f_{i,0}|\mu_0, \boldsymbol{\theta}] \sim GP(\mu_0, \Sigma)$$

$$p[\mu_0| \ \boldsymbol{\theta}] \sim GP(0, \mathrm{C})$$

$$p[\boldsymbol{V_\epsilon}, \boldsymbol{\theta}] \qquad (6)$$

References

Beddington, J. R., C. A. Free, and J. H. Lawton. (1975). Dynamic complexity in predator-prey models framed in difference equations. *Nature*, 225, 58-60.

Bjornstad, O.N., and Grenfell, B.T. (2001). Noisy Clockwork: time series analysis of population fluctuations in animals, *Science, 293*, 638-643.

Blum, M. and Riedmiller, M.A. (2013). Optimization of Gaussian process hyperparameters using Rprop, *In European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Belgium, 24-26.

Buzug, Th. and Pfister, G. (1992). Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local behavior of strange attractors, *Phys. Rev. A, 45:10,* 7073-7084.

Carpenter, S.R. (2003). Regime Shifts in Lake Ecosystems: Pattern and Variation, *Excellence in Ecology Series, 15*, Ecology Institute, Oldendorf/Luhe, Germany.

Chen, B. and Tong, H. (1992). On consistent nonparametric order determination and chaos, *J. Roy. Stat. Soc. B (Methodological), 54:2*, 427-449.

Cressie, N. (1993). *Statistics for spatial data* (revised). Wiley

Cushing, J.M., Levarge, S., Chitnis, N., and Henson, S.M. (2004). Some discrete competition models and the competitive exclusion principle, *J. Diff. Eq. Appl., 10:13-15*.

Desharnais, R.A., Constantino, R.F., Cushing, J.M., Henson, S.M., and Dennis, B. (2001). Chaos and population control of insect outbreaks, *Ecol. Lett., 4:* 229-235.

Deyle E.R., Sugihara G. (2011). Generalized theorems for nonlinear state space reconstruction *PLoS ONE, 6:3, e18295*.

Deyle, E.R., Fogarty, M, Hsieh, C.-h., Kaufmann, L., MacCall, A.D., Munch, S.B., Perretti,C., Ye, H., and Sugihara, G. (2013). Predicting climate effects on Pacific sardine, *PNAS*, *110: 6430-6435*

Doebeli, M. and de Jong, G., 1999. Genetic variability in sensitivity to population density affects the dynamics of simple ecological models. *TPB*, 55, 37-52.

Ellner, S. and Turchin, P. (1993). Chaos in a “noisy” world: new methods and evidence from time series analysis, *Am. Nat., 145:* 343-375.

Gerber, L.R., Karieva, P.M., and Bascompte, J. (2002). The influence of life history attributes and fishing pressure on the efficacy of marine reserves, *J. Biol. Cons., 106*: 11-18.

Ginzurg, L.R. and Taneyhill, D.E. (1994). Population cycles of forest Lepidoptera: a maternal effect hypothesis, *J. Anim. Ecol., 63,* 79-92.

Grenfell, B.T., Brornstad, O.N., and Kappey, J. (2001). Travelling waves and spatial hierarchies in measles epidemics, *Nature, 414*, 716-723.

Halstead, B.J., Wylie, G.D., Coates, P.S., Valcarcel, P., and Casazza, M.L. (2012). Bayesian shared frailty models for regional inference about wildlife survival, *Animal Conservation, 15*, 117–124.

Hans, C. (2009). Bayesian LASSO regression, *Biometrika*, 1-11.

Hare, S.R. and Mantua, N.J. (2000). Empirical evidence for North Pacific regime shifts in 1977 and 1989, *Progress in Oceanography, 47*, 103-145.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrics*, 57: 97-109.

Hastings, A., Hom, C.L., Ellner, S., Turchin, P., and Godfray, H.C.J. (1993). Chaos in ecology: is mother nature a strange attractor?, *Annual Review of Ecology and Systematics,* 24, 1-33.

Hsieh, C-h., Anderson, C., and Sugihara, G. (2008). Extending nonlinear analysis to short ecological time series, *The American Naturalist,* 171: 71-80.

Ives, A.R. and Dakos, V. (2012). Detecting dynamical changes in nonlinear time series using locally linear state-space models, *Ecosphere,* 3(6):58.

Kannathal, N., Choo, M.L., Acharya, U.R., and Sadasivan, P.K. (2005). Entropies for detection of epilepsy is EEG, *Computer Methods and Programs in Biomedicine, 80,* 187-194.

Lee, T.H., Shiba, S., and Wood, R.C. (1999). Integrated management Systems: A Practical Approach to Transforming Organizations. Jon Wiley and Sons, Inc.

Mayfield, E.S. and Mizrach, B. (1992). On determining the dimension of real-time stock-price data, *Journal of Business and Economic Statistics, 10:3,* 367- 374.

Ming-Hui, C. and Shao, Q.-M., and Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag, New York.

Munch, S.B., Kottas, A. and Mangel, M. (2005). Bayesian nonparametric analysis of stock–recruitment relationships, *Can. J. Fish. Aquat. Sci. 62*, 1808–1821.

Neal, R.M. (1996). *Bayesian Learning for Neural Networks*, Springer-Verlag, New York.

Nocedal, J. and Wright, S. (1999). *Numerical Optimization*. Springer-Verlag, New York.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction, *Journal of the Royal Statistical Society B, 40,* 1–42. (with discussion).

Patterson, T.A., Thomas L., Wilcox, C., Ovaskainen, O., and Matthiopoulos, J. (2008). State-space models of individual animal movement, *Trends in Ecology and Evolution, 32:2,* 87-94.

Perretti, C.T., Munch, S.B., and Sugihara, G. (2013). Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data, *PNAS*, *110:5253-5257*

Rasmussen, C.E. and Williams, C. (2006). *Gaussian processes for machine learning*. The MIT Press.

Riedmeiller, M. and Braun, H. (1993). A direct adaptive method for faster backpropagation: the RPROP algorithm, *In Proceedings of the IEEE International Conference on Neural Networks,* San Francisco, CA.

Royle, J.A., and Dorazio, R.M. (2008). *Hierarchical Modeling and Inference in Ecology: the analysis of data from populations, metapopulations and communities*. Elsevier/Academic Press, London, UK.

Sacks, J. and Ylvisaker, D. (1966). Designs for regression problems with correlated errors, *The Annals of Mathematical Statistics, 37:1*, 66-89

Schaffer, W.M. (1985). Order and chaos in ecological systems, *Ecology, 66,* 93-106.

Shi, J.Q., Murray-Smith, R., Titterington, D.M. (2005). Hierarchical Gaussian process mixtures for regression, *Journal of Statistics and Computing, 15*, 31-41.

Stark, J. (1999). Delay embeddings for forced systems. I. Deterministic forcing, *Journal of Nonlinear Science, 9*, 255-332.

Stark, J., Broomhead, D.S., Davies, M.E., and Huke, J. (2003). Delay embeddings for forced systems. II. Stochastic forcing, *Journal of Nonlinear Science, 13*: 519-577

Sugeno, M. and Munch, S.B. (2013). Semiparametric Bayesian method for detecting Allee effects, *JEcology, 94*: 1196-1204.

Sugihara, G., Grenfell, B., May, R.M., Chesson, P., Platt, H.M., and Williamson, M. (1990). Distinguishing error from chaos in ecological time series, *Philosophical Transactions: Biological Sciences, 330:235-251*

Sugihara, G. (1994). Nonlinear forecasting for the classification of natural time series, *Philosophical Transactions: Physical Sciences and Engineering, 348:,* 477-495.

Summers, D., Cranford, J. G., and Healey, B. P. (2000). Chaos in periodically forced discrete-time ecosystem models. Chaos, Solitons & Fractals, 11, 2331-2342.

Takens, F. (1981). Detecting strange attractors in turbulence. In D.A. Rand and L.-S. Young, *Dynamical Systems and Turbulence, Lecture Notes in Mathematics,* Springer-Verlag.

Thorson, J.T., Ono, K., and Munch, S.B. (2014). A Bayesian approach to identifying and compensating for model misspecification in population models, *Ecology, 95*: 329-341.

Tierney, L. and Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, *81*: 82–86.

Turchin, P. (1990). Rarity of density dependence or population regulation with lags?. *Nature*, 344: 660-663.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models*, Springer-Verlag: New York

Wikle, C.K. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes, *Ecology, 84:* 1382-1394.

Wood, S.N. and Thomas, M.B., 1999. Super–sensitivity to structure in biological models. *Proceedings of the Royal Society of London B: Biological Sciences*, 266:.565-570.